Investigating the Effect of Trust Manipulations on Affect over Time in Human-Human versus Human-Robot Interactions

Sarah A. Jessup
Air Force Research Laboratory
Wright-Patterson AFB, OH
sarah.jessup.ctr@us.af.mil

Anthony M. Gibson
Consortium of Universities
Dayton, OH
anthony.gibson.9.ctr@us.af.mil

August Capiola
Air Force Research Laboratory
Wright-Patterson AFB, OH
august.capiola.1@us.af.mil

Gene M. Alarcon
Air Force Research Laboratory
Wright-Patterson AFB, OH
gene.alarcon.1@us.af.mil

Morgan Borders
Wright State University
Dayton, OH
morgan.borders@wright.edu

Abstract

The current study explored the influence of trust and distrust behaviors on affect over time. We examined the differences in affect when participants (N=97) were paired with a human or a robot while playing a modified version of the investor game. Results indicated that there were no differences in affect between partner types when the partner performed a trustful behavior. When the partner performed a distrustful behavior, positive affect was higher for human partners than for robot partners. When robot partners performed a distrustful behavior, negative affect had a steeper incline compared to human partners. These findings suggest that people are more sensitive to distrust behaviors that are performed by a robot over a human.

1. Introduction

Automation is becoming ever more intertwined with our day-to-day experiences. Engineers, researchers, roboticists, and designers are working together to examine how to improve human experiences with automation. The goals of these interactions vary depending on the context. For example, when interacting with self-driving cars, design goals are user safety and preserving calibrated user trust in the car's automation. In comparison, when interacting with a social robot, the design goals may be user entertainment and efficient information sharing between the user and the robot. Researchers are interested in the factors that influence these human-automation (H-A) interactions. The more industry knows about factors that influence these interactions, the more they can improve automation so that it is operating at an optimal level for each user and context.

One of these factors that influence human-automation interaction is trust. Users are less likely to rely on automated systems and robots that they do not trust [1]. Additionally, research has shown that when teammates trust a robot, performance on tasks is better compared to when teammates distrust robots [2; 3].

Another factor that has influenced human-automation interactions is affect (e.g., [4]). Similar to when people interact with one another, the emotions experienced while engaging with automation and robots can influence that interaction in both positive and negative ways. Further, people may experience different affective responses depending on whether they trust or distrust an automated referent [1]. In this paper, we discuss the roles of trust and affect in human-human (H-H) versus H-A interactions. Our main goal is to examine the effect of trust manipulations on affect.

1.1. Trust

Trust is defined as a willingness to be vulnerable to another with the expectation of a positive outcome [3]. A *trustor* is the person engaging in trusting intentions or actions, and a *trustee* is the referent or object of trust. Mayer and colleagues [5] proposed a model of the trust process that explicates trust from its antecedents, namely the trustor's propensity to trust (i.e., a general tendency to trust others) and the trustor's perceived trustworthiness of the trustee (i.e., characteristics of trustees that influence how trustworthy they appear to the trustor). These antecedents influence the

URI: https://hdl.handle.net/10125/63807
978-0-9981331-3-3

trustor's willingness to be vulnerable (i.e., trust intention), which leads to reliance behaviors (i.e., the behavioral outcomes attributed to a trust intention). There is a considerable amount of empirical research that has examined the trust process in H-H interactions. However, there is comparatively less research that has examined differences in the trust process in H-H versus H-A interactions.

Although people apply social norms to human-automation interactions (e.g., [6]), there may be differences between how humans trust one another and how humans trust automation. Madhaven and Weigmann [7] described the similarities and differences in these two types of interactions. People commit the fundamental attribution error when engaging with both humans and automation. That is, whether the referent is a person or an automated system, people attribute undesirable behaviors to the referent's personality (or the entire system in the case of automation) instead of situational circumstances. However, differences between H-H and H-A interactions reside in the trust process itself, mainly trust building and decaying over time. In interpersonal interactions, people are more cautious in the beginning of a relationship: it takes longer to build a relationship and to establish trust between two people. In H-A interactions, there is an automation bias (e.g., [8]) where humans trust automation more than humans in initial interactions [7]. This is also similar to a bias known as perfect automation schema [9]. Perfect automation schema is the belief that automated systems perform without errors and have better, more reliable performance than humans. Thus, people are more forgiving of humans when they make a mistake or perform contrary to their expectations. It is easier for people to rationalize humans' actions. In comparison, if a system or robot makes a mistake, people's trust in that system dramatically decreases. Compared to H-A interactions, in H-H interactions, it takes longer for trust to decrease and a shorter time to recover over the course of a relationship [7]. In both H-H and H-A interactions, changes in perceived trustworthiness might influence self-reported affect.

1.2. Positive and Negative Affect

Emotions are "organized responses, crossing the boundaries of many psychological subsystems, including the physiological, cognitive, motivational, and experiential systems. Emotions typically arise in response to an event, either internal or external, that has a positively or negatively valanced meaning for the individual" [10]. Emotions vary on two dimensions: arousal and valence. Arousal, or intensity of the emotion, ranges from low to high. Valence ranges from positive to negative. Positively valanced emotions (i.e., positive affect) are described as happy, enthusiastic, and alert. Examples of negatively valanced emotions (i.e., negative affect) are anger, fear, and disgust [11]. People use both affect and cognition to help interpret situations and aid in decision-making [12]. For example, if people feel emotionally connected to robots, their perceived trustworthiness of the robot may increase [2; 13], demonstrating that positive affect (PA) may affect judgments towards robots. Thus, interplay of affect and trust is an important consideration for trust research.

1.2.1. Positive Affect and Trust

Researchers have found that participants who experienced PA (e.g., happiness) rated referents as more trustworthy compared to participants who experienced
negative affect (NA), namely anger [14]. Furthermore, affect influenced trust only when participants rated someone who was unfamiliar to them, such as an acquaintance compared to a familiar person. Similarly,
[15] primed participants with PA or NA prior to an experimental task. They found that participants who were assigned to the PA condition reported feeling PA prior to an automated convoy task and reported higher trust in an automated decision aid during the task. However, these effects were only demonstrated in the first session. The subsequent two sessions that participants completed did not show this effect. It appears that affect only influences initial trust, or trust from the most recent transaction. As people acquire more information about the referent, other factors become significant predictors of trust other than affect. Lount [16] reported that when participants were provided with information on how trustworthy their partner was via self-report scores in a trust game, there was an interaction between affect and how much money the participants sent to their partners. When participants were in a positive affective state, they sent more money to trustworthy partners, compared to participants who experience neutral affect, in which there were no differences in how much money they sent to their partners.

These studies demonstrate that PA influences trust in a relationship when people have little information about their partners. The current study investigates a different directional hypothesis—the role of trust and distrust manipulations on affective

responses. Though the relationship between trust and affect has been examined in the aforementioned research [2; 3; 13; 14; 15; 16], the current study wishes to narrow focus on the effect of trust manipulations on affect, and how biases towards humans and robots affect these relationships.

1.2.2. Negative Affect and Trust

In our review of the published literature on emotions and trust in the context of game theory, anger was the most discussed and prevalent negative affect emotion. In particular, anger can be triggered by low offers from the trustor in Trust and Ultimatum games [17]. Games such as the Trust, Ultimatum, and Investor/Dictator games are games in which usually two people exchange money between one another to study fairness, trust, and self-interest [17]. The first player (e.g., trustor, prosper, or investor), is given money from the experimenter at the start of the session and told he/she can split it with the second player (e.g., trustee, dictator, or responder). Depending on the game, the session can last one or multiple rounds. The trustee can choose to accept the offer or reject it and the game ends (Ultimatum and Investor/Dictator game), or the money is tripled each time it is passed to each player and the players have the option to stop the game at any point and take the entire earnings (Trust game). Hewig and colleagues [18] studied how participants felt after playing both the Ultimatum and Dictator games. Results indicated that as unfair offers increased, participants reported more negative emotions.

Pillutla and Murnighan [19] examined the effect participants' anger had on their rejections when their partner in the Ultimatum game made an unfair offer. Results indicated that anger and unfairness were significantly, positively correlated, such that as unfair offers increased, so did anger. These two studies examined how NA can influence how participants behave while playing games designed to study trust. Anger in particular was positively correlated with higher rejection rates. One explanation is that participants felt as though they were being treated unfairly. However, these studies only compared H-H dyads and did not examine H-A pairs.

In an effort to study the role of biases in H-H pairs compared to H-A pairs, researchers compared how humans responded both behaviorally and physiologically when playing the Ultimatum game with both human and computer partners [20]. When participants received a low, unfair offer of money from their partners, participants were more likely to reject those offers in the H-H condition when compared to the human-computer condition. Also, participants had high emotional arousal as measured by skin conductivity when they were offered low offers from another human. Conversely, when participants played with a computer partner, there were no differences in emotional arousal. These results could be because people perceived the computer as fair and thus failed to experience negative emotions. The current study seeks to investigate these effects over time. Specifically, we explored the effects of trust manipulations on affect over time, and how this relationship is moderated by characteristics of the partner (human vs automation).

1.3. The Current Study

The aim of the current study is to examine the change in affect over time in unfamiliar H-H and human-robot (H-R) interactions. We examined self-reported state affect changes when participants' partners display trustful and distrustful behaviors. Before making directional hypotheses, the task should be explained so there is more context for each hypothesis.

2. Method

2.1. Participants

Participants were 97 adults recruited from a Midwestern college. Participants were randomly distributed among the four experimental conditions: Trust-Human (n = 23), Distrust-Human (n = 22), Trust- Robot (n = 25), or Distrust-Robot (n = 27). Ages ranged from 18-41 years (M = 22.82 years, SD = 4.68 years). Most (59%) were female and white (41%). Participants were recruited from the Introduction to Psychology participant pool, flyers, email, and word of mouth. Participants received compensation in the form of a $30 gift card, as well as cash payment for all money earned during the task. The study was overseen by the institutional review board.

2.2. Task

The task played in the current study is called Checkmate [21]. It is a computer game played between two players. Checkmate is a modified version of the investor/dictator game [22]. In the current study, the participant was assigned the role of the "Banker" (investor in the investment/dictator game) and a robot or confederate played the role of the "Runner" (dictator in the investment/dictator

game). The role of the Banker was to loan money to the Runner over the course of five rounds. The role of the Runner was to collect boxes in a virtual maze over the course of five rounds. The number of boxes collected by the Runner reflected performance. The initial amount of money in the Banker's virtual account was set at $50. The Bankers loaned money to the Runner each round in anticipation of earning interest on their investment. Each round the Banker chose to loan one of three amounts to the Runner: small ($1-$7), medium ($4-$10), or large ($7-$13). Based on their selections, a pre-determined algorithm specified the exact dollar amount that would be sent to the Runner.

The Runner chose a risk level for the purpose of potentially increasing the initial loan amount. The risk levels were low (75-150%), moderate (50-200%), and high (0-300%). The Runner could earn more money by choosing a higher risk level, but the Runner risked not earning any money at all if his performance was poor. If the Runner decided to err on the side of caution and chose a low risk level, the maximum amount of money the Runner lost was 25% without collecting any boxes or gained 50% by performing well.

At the beginning of the round, the Runner chose a risk level. The Runner then promised to return the initial loan and 50% of the earnings to the Banker. The Banker was notified via a pop-up message which risk level the Runner selected, as well as how much of the invested money the Runner promised to return. At this point in the round, the Banker selected an amount to loan to the Runner. Money was then transferred into the Runner's virtual wallet. The maze-running task began, and the Banker was able to watch a top-down video of the Runner's progress. The Runner was allotted two minutes to collect as many boxes as possible. After the maze-running task was over, the Runner then decided how much money to return to the Banker. The Banker received a pop-up message of the exact amount of money the Runner decided to return. If the amount returned was within the range of what the Runner had promised, then the Banker could assume that the Runner was trustworthy. However, if the return amount was lower than promised, then the Banker might assume that 1) the Runner may have not earned enough money to return and keep their promise, or 2) the Runner is playing unfairly by keeping more money for themselves, which could signal that the Runner is distrustful.

The steps outlined above were repeated over six rounds, which we coded as zero to five. Participants were informed that the amount of money the Banker had in his/her virtual bank at the end of the session belonged to the Banker, and the earnings were paid out in the form of cash, rounded up to the nearest quarter.

2.3. Manipulations

Typically, Checkmate [21] is played between two people. For this study, the participant was always the Banker, and the Runner was either a Nao robot (see Figure 1) or a male confederate. The Runner's risk level in the game was set to medium-risk for every round. All the Runner's data, including maze performance and returning of investment to Banker, was prerecorded. This level of control allowed a focus on the way that participants trusted their partner. However, participants were led to believe they were playing in real time with either the robot or the human. Additionally, participants were randomly assigned to one of two experimental conditions: trust or distrust. In the trust condition, the Runner always returned the amount of money that was promised for rounds 0-5. In the distrust condition, the Runner returned less money than he promised for rounds 3 and 4.

2.4. Measures

2.4.1. Affect.

State affect was measured using the shortened 10-item Positive and Negative Affect Schedule (PANAS; [11]). Participants were instructed to indicate the extent they felt in the present moment using a 5-point response scale (1 = *very slightly or not at all*, 5 = *extremely*). Scores were computed by averaging responses for the positive and negative affective words separately so that each participant had an independent average score for both PA and NA for rounds 0-5.

Positive affect (PA) items included Interested, Excited, Enthusiastic, Alert, and Determined. Scale reliabilities are as follows: Round 0 (α = .87); Round 1 (α = .90); Round 2 (α = .88); Round 3 (α = .88); Round 4 (α = .89); Round 5 (α = .89).

Negative affect (NA) items included Distressed, Upset, Irritable, Nervous, and Jittery. Scale reliabilities are as follows: Round 0 (α = .76); Round 1 (α = .61); Round 2 (α = .71); Round 3 (α = .72); Round 4 (α = .74); Round 5 (α = .76).

2.4.2. Time.

Time was classified at each round. The practice round was coded as Time 0 and the five subsequent rounds were coded as Time 1-5.

Figure 1. NAO robot; partner in robot condition.

2.5. Procedure

Participants were run individually in a two-room laboratory. First, they were introduced to their partner (robot or confederate). In the robot condition, the robot was located in the back room of the computer lab. The participants were told they were going to meet the other participant for the study, and then walked into the back room to meet the robot. The experimenter tapped the robot on the head, which initiated the following speech and behavior. The robot stood up and became animated and said the following, "Thanks for waking me up [experimenter's name]. Hi, I'm Rufus. It's nice to meet you. Time to get to work." Then the robot returned to the crouching position. In the human condition, participants were introduced to each other once they entered the lab together and then seated in separate rooms.

After providing informed consent, participants completed demographic surveys, then completed an endowment earning task, which consisted of five, medium-difficulty, multiple choice math problems. The purpose of this task was to make participants feel like they earned the money. Because the money in the task was in a virtual bank, we wanted to make this connection as salient as possible. Participants were told that based on their performance they would earn money towards the main task if they answered at least three out of five of the questions correctly. However, all participants earned $50 regardless of their performance in order to ensure experimental control. After the math task, in the robot condition, the experimenter read a backstory on Rufus aloud to participants, "The military currently integrates automation into dangerous scenarios alongside humans. Automation is useful in high-risk scenarios, such as disabling explosive devices, navigating unmanned aerial vehicles (UAVs), and carrying heavy equipment. However, automation is expensive and takes time to develop. As such, the military is testing automated robots containing self- preservation algorithms. This means the military is creating robots that should be able to make decisions to protect themselves, as well as other humans around them. If a situation is too dangerous, the robot should take proper precautions to minimize damages to itself. The current study uses the same algorithms to aid the robot's decision-making when teamed with another human in a maze-running task. Keep in mind that Rufus the robot may act self-interested, meaning he may prioritize himself over you."

Next, participants completed training on Checkmate, then played a practice round of Checkmate with their partner. Participants were told prior to coming in that they were randomly selected to play the Banker for the real session of five rounds and their partner was selected to play the Runner. Following practice (Time 0), participants completed the state affect questionnaire. Each round lasted approximately three to five minutes. Following each round (Time 1-5), participants were asked to complete the state affect survey. After the competition of the fifth round, participants were debriefed and paid for their time with a $30 gift card. The money in their virtual wallet was paid to them in cash.

2.6. Research Design and Analysis

We tested changes in self-reported affect over time across the Condition (Trust vs Distrust) and Partner (Human vs Robot) factors using growth curve models
[23] in the nlme package in R [24; 25]. Growth curve models have benefits over repeated measures ANOVA (e.g., more relaxed model assumptions, ability to handle missing data, ability to model individual growth patterns). In general, there are two levels to growth models. Level-1 variables correspond to time-level variables (e.g., time, time-variant covariates), whereas Level-2 variables occur at the person level (e.g., time- invariant covariates). In the current models, we denoted three random effects (i.e., intercept variance, a quadratic time term variance, and a cubic time term variance). Random slope variance (i.e., random quadratic and cubic effects) allow each person to have his or her unique growth estimate. Then, we predicted that individual growth curve with person-level variables (i.e., partner type and condition).

Overall, we expected a linear change over time for the trust condition and a cubic change over time for the distrust condition. In general, polynomial terms model deviations from the typical linear

regression. The number of "bends" modeled in the growth curve can be calculated by subtracting one from the polynomial term. For example, a cubic term allows the slope to bend twice. In the distrust condition specifically, we expected that PA would increase for the first three rounds, decrease after the two distrust behaviors, and then increase after the final trust behavior (i.e., a cubic slope). We expected the opposite pattern for the NA model (i.e., an initial decrease in NA, a sharp increase in NA following the two distrust behaviors, and finally a decrease in NA after the final trusting behavior). We also predicted that NA would rise steeper following a distrust behavior when the partner type was human, given that people ascribe their feelings of anger and spite to humans more than automation [20]. Thus, our hypotheses are as follows:

RQ: Are there differences in PA and NA for partner type for Time 0-2?

H1: In both conditions, PA will increase for Time 0-2.

H2: In the distrust condition, PA will decrease for Time 3 and Time 4, and increase in Time 5.

H3: In the distrust condition, PA will decrease more over time when the partner is a human compared to a robot.

H4: In the distrust condition, PA will be higher for robot than human

H5: In both conditions, NA will decrease for Time 0-2.

H6: In the distrust condition, NA will increase for Time 3 and Time 4, and decrease in Time 5.

H7: In the distrust condition, NA will increase more if the partner is a human compared to a robot.

H8: In the distrust condition, NA will be higher for human than robot.

3. Results

3.1. Positive Affect

First, we determined whether the intraclass correlation (ICC) was large enough to allow each person to have a unique initial PA score. We found an ICC score of .64, which supports a free intercept model (i.e., allowing each participant to have a unique starting PA score). We observed significant estimates for time (B = 0.46, $t(473)$ = 5.67, p < .01), the quadratic term for the distrust condition (B = -0.09, $t(473)$ = -2.98, p < .01), and the cubic term for the distrust condition (B = 0.01, $t(473)$ = 2.49, p < .05). Stated simply, participants across both trust and distrust conditions showed a significant increase in PA across the first three trials (Time 0-2), supporting Hypothesis 1. Participants assigned to the distrust condition showed a significant decrease in PA following the distrust behaviors, and then a significant increase in PA following the final trust behavior (see Figure 2). Hypothesis 2 was supported. We also found evidence of auto-correlated errors, $\Delta\chi^2(1)$= 11.37, p < .01, so we included this term in the model. This accounts for the measurement errors of proximal time points having stronger correlations with each other than measurements more distally spaced in time. We found no evidence of violation of the homoscedasticity assumption, so we excluded it from the model. We found no significant differences in quadratic or cubic time between partner types (see Table 1). Hypotheses 3 and 4 were not supported.

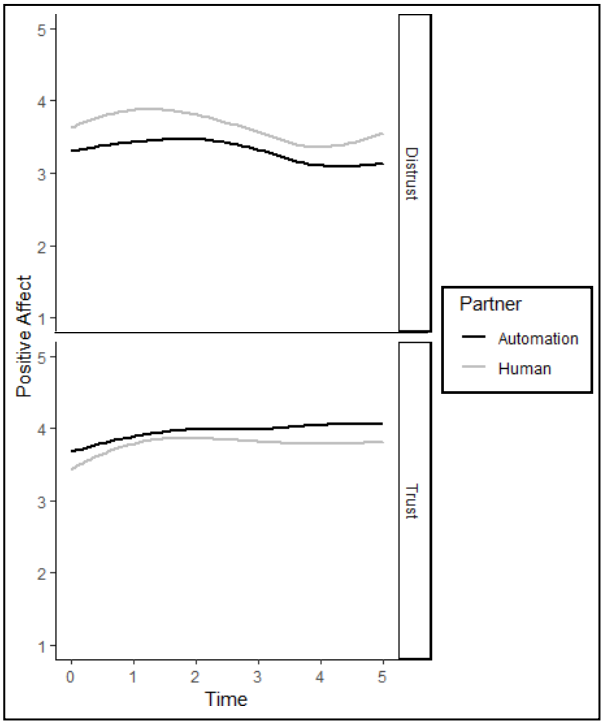

Figure 2. Change in positive affect over time in the distrust condition (top) and the trust condition.

3.2. Negative Affect

We followed the same steps for testing differences in NA as described in the PA section above. We observed an intraclass correlation (ICC) of .55, so we allowed the intercepts to vary across people. We observed a significant estimate for time (B = -0.62, $t(473)$ = -8.57, p < .01), a significant quadratic term

for the distrust condition (B = 0.13, $t(473)$ = 4.48, p < .01), and a significant cubic term for the distrust condition (B = - 0.02, $t(473)$ = -4.14, p < .01). Participants across both conditions showed a decrease in NA across the first three trials. Hypothesis 5 was supported. Then, those assigned to the distrust condition showed a significant increase in NA following the distrust behaviors, and a significant decrease in NA following the final trust behavior (see Figure 3). Hypothesis 6 was supported. We found evidence of violation to the assumption of homoscedasticity of the errors, $\Delta\chi^2(1) = 9.18$, $p < .01$, so we included this term in the model [23]. We observed no significant differences in human and robot partners on decreases in NA across the first three time points. In the distrust condition, the increase in NA after the distrust behaviors was stronger for the robot condition, γ = 0.12, t(468) = 2.03, p < .05. Thus, Hypothesis 7 was not supported. The decrease in NA following the final trust behavior was also steeper for the robot partner type, γ =-0.02, t(468) = -2.21, p < .05. Hypothesis 8 was not supported. Note that these findings were the opposite of the predicted pattern.

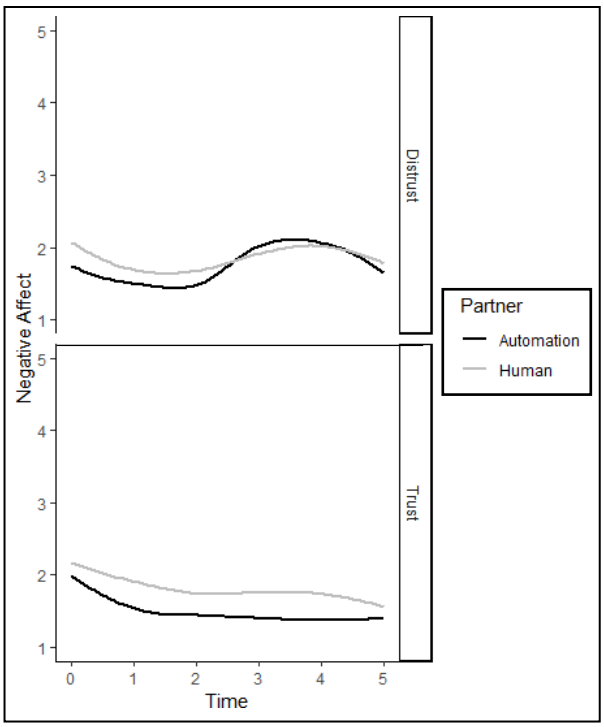

Figure 3. Change in NA over time in the distrust condition (top) and the trust condition (bottom).

Table 1
Positive Affect Changes Over Time

	Estimate	*SE*	*df*	*t*-value	*p*-value
(Intercept)	3.42	0.20	468	17.48	.00
Time	0.61	0.12	468	4.95	.00
Time^2	-0.24	0.07	468	-3.59	.00
Time^3	0.03	0.01	468	2.87	.00
Partner	0.24	0.27	93	0.88	.38
Trust	0.21	0.27	93	0.75	.45
Time:Partner	-0.29	0.17	468	-1.73	.09
Time^2:Trust	-0.12	0.05	468	-2.49	.01
Partner:Time^2	0.14	0.09	468	1.58	.12
Partner:Trust	-0.56	0.37	93	-1.49	.14
Trust: Time^3	0.02	0.01	468	2.25	.02
Partner: Time^3	-0.02	0.01	468	-1.33	.19
Partner:Time^2:Trust	0.04	0.06	468	0.55	.58
Partner:Trust: Time^3	-0.01	0.01	468	-0.67	.51

Note. Time = linear change. Time2 = quadratic change. Time3 = cubic change. Partner = Human vs. Robot. Robot was the reference group. Trust = Distrust was the reference group.

Table 2
Negative Affect Changes Over Time

	Estimate	*SE*	*df*	*t*-value	*p*-value
(Intercept)	2.25	0.13	468.00	16.99	.00
Time	-0.62	0.11	468.00	-5.39	.00
$Time^2$	0.24	0.06	468.00	4.01	.00
$Time^3$	-0.03	0.01	468.00	-3.55	.00
Partner	-0.26	0.18	93.00	-1.42	.16
Trust	-0.23	0.18	93.00	-1.31	.19
Time:Partner	-0.01	0.16	468.00	-0.07	.94
$Time^2$:Trust	0.08	0.04	468.00	1.83	.07
Partner:$Time^2$	-0.04	0.08	468.00	-0.47	.64
Partner:Trust	-0.01	0.24	93.00	-0.04	.96
Trust: $Time^3$	-0.01	0.01	468.00	-1.52	.13
Partner: $Time^3$	0.01	0.01	468.00	0.80	.42
Partner:$Time^2$:Trust	0.12	0.06	468.00	2.03	.04
Partner:Trust: $Time^3$	-0.02	0.01	468.00	-2.21	.03

Note. Time = linear change. Time2 = quadratic change. Time3 = cubic change. Partner = Human vs. Robot. Robot was the reference group. Trust = Distrust was the reference group.

4. Discussion

Overall, we expected participants to report increased PA, and decreased NA, when participants experienced a trust behavior. Additionally, we predicted that when participants were paired with a human partner, participants would have steeper changes in affect following a distrust behavior compared to a robot partner. As hypothesized, we found that PA had a linear relationship with time, such that PA increased for the first three time points in both conditions, regardless of partner type. Just as PA increased for the first three rounds in both conditions, NA decreased for Time 0-2, demonstrating an expected negative relationship between PA and NA. This is understandable, as the first three time points were all trust behaviors. For Time 3 and Time 4 in the distrust condition, PA decreased and NA increased when the participants received less money back than promised from their partners. During the final round (Time 5), when the partner once again returned the amount of money that was promised, PA increased and NA decreased. However, contrary to our hypotheses, the change in PA in the distrust condition, depending on partner type, was non-significant. There were no differences in PA when participants' partner was a human or a robot. There was a significant change in NA in the distrust condition, although it was in the opposite direction we hypothesized. Specifically, NA increased more when the partner was a robot compared to a human, and PA was higher when partner type was a human compared to a robot. These results contradict what past researchers have found [20]. One reason for this could be that the type of automation that used was a computer [20], and we used a robot. As automation becomes more anthropomorphized, people ascribe more human-like qualities to the referent [6, 26]. As such, people may attribute will and autonomy to robots more than they do computers. Additionally, these differences may have been due to differences in the task. Future research should compare various types of automation to a human partner in a variety of tasks to examine the effects of anthropomorphism on affect.

Pulling from the social science literature, another reason that these results could have occurred is due to person-positivity bias [27]. This means that people generally believe the best in people and are optimistic about others' intentions. Similarly, the mere-exposure effect [28], sometimes referred to as the familiarity principle, posits that people rate others and objects more positively when they are familiar with them. In this study, the partner type was either a human or a robot. As the participants were human, they were more familiar with the human partner compared to a robot partner. Therefore, one reason that the participants experienced less PA and more NA when their partner was a robot is because, presumably, they have had limited exposure to anthropomorphized robots.

Finally, the current findings align with prior research on automation bias [8; 29]. If participants perceived the robot partner as being perfectly reliable, they may have experienced increased NA following a distrust behavior due to violations of this heuristic. It is noteworthy, however, that NA decreased more for the robot partner compared to the human partner for the trust behavior following the distrust behaviors. We would expect that violations to the perfect automation schema would result in a slower decrease in NA in trust recovery compared to a human partner.

A limitation of this study, however, was that it only contained a total of six measurement time points. In order to gain a better understanding of how trust behaviors influence affect over time, more instances of trust and distrust should be included. Specifically, we may have observed a different pattern in affect amongst the trust recovery process with more time points included after time point four. However, to our knowledge, this is one of the few studies that has measured affect over several time points, whereas most studies are cross-sectional.

Another limitation of the current study is the sample size. Although researchers are unsure of the exact sample size required for growth curve models [30], a larger sample size may be needed given our limited number of measurements and the cubic nature of the change for those assigned to the distrust conditions. We should note, however, that the number of longitudinal studies on comparing H-H and H-R trust has been minimal.

A third limitation concerns our sample of participants. This was a convenience sample of mostly college students from a Midwestern university. Previous researchers have demonstrated that while recruiting participants using various methods such as crowd- soured websites like Amazon Mechanical Turk or social media postings on platforms such as Twitter or Reddit result in more diverse samples compared to college student samples, results from an in-lab behavioral study with college students had almost identical results when it was adapted for a computer and administered online to participants from crowd-sourced and social media outlets [31]. However, our research may not generalize to H-R teams using people in HRI in the real world.

We used a shortened version of the PANAS [11], and some items may have been ambiguous in the current context. For example, some item stems (e.g., Distressed, Nervous, Jittery in the NA scale; Excited, Enthusiastic within the PA scale) may have been inappropriate for this context, because the task itself did not lend itself to evoke these emotions. Future research may benefit from using all 20 items or selecting affect items that are more likely to be experienced during the task.

Finally, we omitted the relationship between affect and actual participant behavior. The statistical models used in the current study were complex, and analyses on categorical behavioral outcomes only add to the complexity. Moreover, the addition of these analyses were outside the scope of this study which focused on the effects of a trust manipulation on affect. Given the practical significance of the effects of PA and NA on perception and behaviors [32]; future research should examine behaviors when comparing affective outcomes across human and robot partners.

This study demonstrated the influence of trust and distrust behaviors on affect over time. This research is important because affect is essential to judgement, decision-making, and reasoning [33]. The implications for this research concern the affective responses that are attributed to trust manipulations. Contrary to our hypotheses, trust violations led to increased NA responses when the human was partnered with a robot. We postulate this may be due to more severe decay of trust when an automated aid fails to perform as expected [7], and this in turn may lead to a negative affective response. Note, however, that this increase in NA may be beneficial, as prior research has found that NA has related to higher attention to specific details (e.g., [34]). However, violations to the perfect automation schema may lead to automation disuse [9]. Future work may consider these affective responses differing between humans and robots. When training teams comprising humans and automated assistants, researchers should note that a loss of trust between each referent may lead to different affective responses, and this trajectory may vary over time and have differing consequences for team-based tasks.

This research demonstrated that when robot partners engage in distrust behaviors, humans experience more NA compared to human partners. When users experience NA, they may be less likely to interact with robots. Designers should increase transparency and make sure that users understand the capabilities of the robot in order to reduce instances of NA and promote successful interactions.

5. References

[1] J.D. Lee, and K.A. See, "Trust in Automation: Designing for Appropriate Reliance", Human Factors, SAGE Publications, U.S.A., 2004, pp. 50-80.

[2] S. You, and L.P. Robert, "Emotional Attachment, Performance, and Viability in Teams Collaborating with Embodied Physical Action (EPA) Robots", Journal of the Association for Information Systems, Association for Information Systems, U.S.A., 2018, pp. 377-407.

[3] S. You, and L.P. Robert, "Trusting Robots in Teams: Examining the Impacts of Trusting Robots on Team Performance and Satisfaction", In Bui, T.X. and Sprague, R.H. (Eds.), Proceedings of the 52th Hawaii International Conference on System Sciences, U.S.A., 2019 pp. 244-253.

[4] A. M. Rosenthal-von der Pütten, N.C. Krämer, L. Hoffmann, S. Sobieraj, and S.C. Eimler, "An Experimental Study on Emotional Reactions Towards a Robot", International Journal of Social Robotics,

Springer, U.S.A., 2013, pp. 17-34.
[5] R.C. Mayer, J.H. Davis, and F.D. Schoorman, “An Integrative Model of Organizational Trust”, Academy of Management Review, Academy of Management, U.S.A., 1995, pp 709-734.
[6] C. Nass, and Y. Moon, “Machines and Mindlessness: Social Responses to Computers”, Journal of Social Issues, Wiley-Blackwell, U.S.A., 2000, pp. 81-103.
[7] P. Madhavan, and D.A. Wiegmann, “Similarities and Differences between Human–human and Human–automation Trust: An Integrative Review”, Theoretical Issues in Ergonomics Science, Taylor & Francis, United Kingdom, 2007, pp. 277-301.
[8] M. Cummings, “Automation Bias in Intelligent Time Critical Decision Support Systems”, 1st Intelligent Systems Technical Conference, In American Institute of Aeronautics and Astronautics, U.S.A., 2004, pp. 1-6.
[9] S.M. Merritt, J.L. Unnerstall, D. Lee, and K. Huber, “Measuring Individual Differences in the Perfect Automation Schema”, Human Factors, SAGE Publications, U.S.A., 2015, pp. 740-753.
[10] P. Salovey, and J.D. Mayer, “Emotional Intelligence”, Imagination, Cognition, and Personality, SAGE Publications, U.S.A., 1990, pp. 185-211.
[11] D. Watson, L.A. Clark, and A. Tellegen, “Development and Validation of Brief Measures of Positive and Negative Affect: The PANAS Scales”, Journal of Personality and Social Psychology, American Psychological Association, U.S.A., 1988, pp. 1063-1070.
[12] J.T. Cacioppo, and G.G. Berntson, “The Affect System: Architecture and Operating Characteristics”, Current Directions in Psychological Science, SAGE Publications, U.S.A., 1999, pp. 133-137.
[13] Siddike, A. Kalam, and Y. Kohda, “Towards a Framework of Trust Determinants in People and Cognitive Assistants Interactions”,. In Proceedings of the 51st Hawaii International Conference on System Sciences. 2018.
[14] J.R. Dunn, and M.E. Schweitzer, “Feeling and Believing: The Influence of Emotion on Trust”, Journal of Personality and Social Psychology, American Psychological Association, U.S.A., 2005, pp. 736-748.
[15] C.K. Stokes, J.B. Lyons, K. Littlejohn, J. Natarian, E. Case, and N. Speranza, “Accounting for the Human in Cyberspace: Effects of Mood on Trust in Automation”, International Symposium on Collaborative Technologies and Systems, Institute of Electrical and Electronics, Unites States, 2010, pp. 180-187.
[16] R.B. Lount Jr., “The Impact of Positive Mood on Trust in Interpersonal and Intergroup Interactions”, Journal of Personality and Social Psychology, American Psychological Association, U.S.A., 2010, pp. 420-433.
[17] Camerer, C.F., Behavioral Game Theory: Experiments in Strategic Interaction, Princeton University Press, U.S.A., 2003.
[18] J. Hewig, N. Kretschmer, R.H. Trippe, H. Hecht, M.G. Coles, C.B. Holroyd, and W.H. Miltner, “Why Humans Deviate from Rational Choice”, Psychophysiology, Wiley- Blackwell, U.S.A., 2011, pp. 507-514.
[19] M.M. Pillutla, and J.K. Murnighan, “Unfairness, Anger, and Spite: Emotional Rejections of Ultimatum Offers”, Organizational Behavior and Human Decision Processes, Elsevier, Netherlands, 1996, pp. 208-224.
[20] M. Van’t Wout, R.S. Kahn, A.G. Sanfey, and A. Aleman, “Affective State and Decision-making in the Ultimatum Game”, Experimental Brain Research, Springer, U.S.A., 2006, pp. 564-568.
[21] G.M. Alarcon, J.B. Lyons, J.C. Christensen, S.L. Klosterman, M.A. Bowers, T.J. Ryan, S.A. Jessup, K.T. Wynne, “The Effect of Propensity to Trust and Perceptions of Trustworthiness on Trust Behaviors in Dyads”, Behavioral Research Methods, Springer, U.S.A., 2018, pp. 1906-1920.
[22] J. Berg, J. Dickhaut, and K. McCabe, “Trust, Reciprocity, and Social History”, Games and Economic Behavior, Elsevier, U.S.A., 1995, pp. 122-142.
[23] P.D. Bliese, and R.E. Ployhart, “Growth Modeling Using Random Coefficient Models: Model Building, Testing, and Illustrations”, Organizational Research Methods, SAGE Publications, U.S.A., 2002, pp. 362-387.
[24] J. Pinheiro, D. Bates, S. DebRoy, D. Sarkar, and R.C. Team, “nlme Linear and Nonlinear Mixed Effects Models” R Package Version 3.1-108, 2013, pp. 111.
[25] R Core Team, R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Austria, 2019.
[26] B.R. Duffy, “Anthropomorphism and the Social Robot”, Robotics and Autonomous Systems, Elsevier, Netherlands, 2003, pp. 177-190.
[27] D.O. Sears, “The Person-positivity Bias”, Journal of Personality and Social Psychology, American Psychological Association, U.S.A., 1983, pp. 233-250.
[28] R.B. Zajonc, “Mere Exposure: A Gateway to the Subliminal”, Current Directions in Psychological Science, Elsevier, Netherlands, 2001, pp. 224-228.
[29] R. Parasuraman, and V. Riley, “Humans and Automation: Use, Misuse, Disuse, Abuse”, Human Factors, SAGE Publications, U.S.A., 1997, pp. 230-253.
[30] P.J. Curran, K. Obeidat, and D. Losardo, “Twelve Frequently Asked Questions about Growth Curve Modeling”, Journal of Cognition and Development, Taylor & Francis, United Kingdom, 2010, pp. 121-136.
[31] K. Casler, L. Bickel, and E. Hackett, “Separate but Equal? A Comparison of Participants and Data Gathered via Amazon’s MTurk, Social Media, and Face-to-Face Behavioral Testing”, Computers in Human Behavior, Elsevier, Netherlands, 2013, pp. 2156-2160.
[32] A.D. Angie, S. Connelly, E.P. Waples, and V. Kligyte, “The Influence of Discrete Emotions on Judgement and Decision-making: A Meta-analytic Review”, Cognition & Emotion, Routledge, United Kingdom, 2011, pp. 1393- 1422.
[33] J. Storbeck, “Negative Affect Promotes Encoding of and Memory for Details at the Expense of the Gist: Affect, Encoding, and False Memories”, Cognition & Emotion, Routledge, United Kingdom, 2013, pp. 800-819.
[34] A.M. Isen, T.E. Shalker, M. Clark, and L. Karp, “Affect, Accessibility of Material in Memory, and Behavior: A Cognitive Loop?”, Journal of Personality and Social Psychology, American Psychological Association, U.S.A., 1978, pp. 1-12.